



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **06215031 A**

(43) Date of publication of application: **05.08.94**

(51) Int. Cl. **G06F 15/40**
G06F 12/00

(21) Application number: 05274684

(22) Date of filing: 02.11.93

(30) Priority: 04.12.92 GB 92 9225424

(71) Applicant: **INTERNATL BUSINESS MACH
CORP <IBM>**

(72) Inventor: **SHARMAN GEOFFREY**

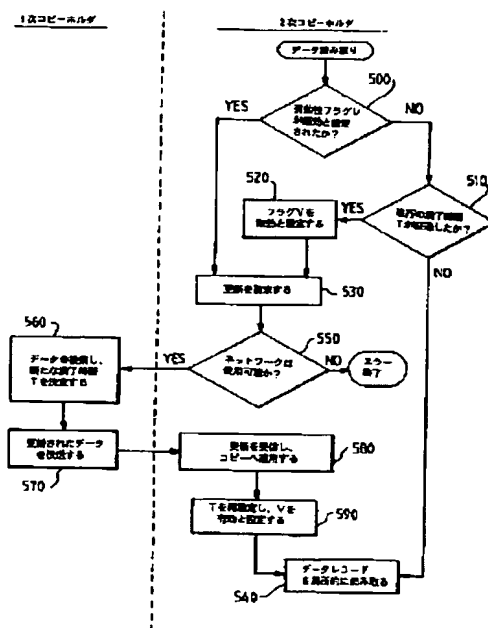
(54) DEVICE AND METHOD FOR DISTRIBUTED DATA PROCESSING

(57) Abstract:

PURPOSE: To provide access to remote data provided with performance and usability matched to the performance and usability obtainable for local data as far as possible.

CONSTITUTION: This distributed data processing method is composed of a step for storing the duplicated secondary copy of a primary data record stored in a first data processor in at least one remote data processor, the step (570) for propagating the updating of the primary data record to the remote data processor, the step for storing the present expiration time corresponding to the respective secondary copies stored in the data processors in the respective remote data processors and the step (510) for checking whether or not the present expiration time passed corresponding to the request of the data access of the data processor.

COPYRIGHT: (C)1994,JPO



(19)日本国特許庁(JP)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平6-215031

(43)公開日 平成6年(1994)8月5日

(51)Int.Cl.⁵G 0 6 F 15/40
12/00

識別記号

5 0 0 M 7218-5L
5 3 3 J 8526-5B

庁内整理番号

F I

技術表示箇所

審査請求 有 請求項の数10 OL (全 13 頁)

(21)出願番号 特願平5-274684

(22)出願日 平成5年(1993)11月2日

(31)優先権主張番号 9 2 2 5 4 2 4 / 2

(32)優先日 1992年12月4日

(33)優先権主張国 イギリス (GB)

(71)出願人 390009531

インターナショナル・ビジネス・マシー
ズ・コーポレーションINTERNATIONAL BUSIN
ESS MACHINES CORPO
RATIONアメリカ合衆国10504、ニューヨーク州
アーモンク (番地なし)

(72)発明者 ジェフリー シャーマン

イギリス国、エスオー22 4エイチエフ
ハンプシャー、ウインチェスター、オリヴ
ァーズ パタリー ガーデنز 12

(74)代理人 弁理士 合田 潔 (外6名)

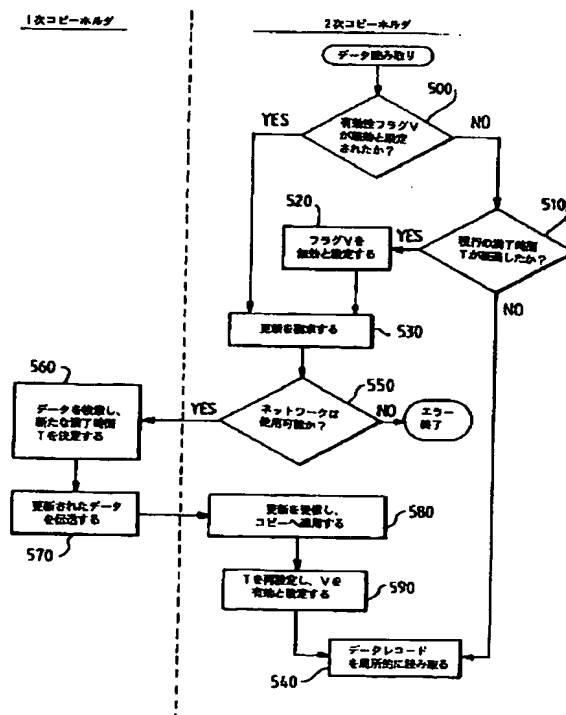
(54)【発明の名称】 分散データ処理装置及び方法

(57)【要約】

【目的】 局所的なデータに対して入手可能なパフォーマンス及び使用可能性と可能な限り一致するパフォーマンス及び使用可能性を備えた遠隔データへのアクセスを提供する。

【構成】 分散データ処理方法は、少なくとも1つの遠隔データプロセッサに第1データプロセッサで記憶される1次データレコードの複写された2次コピーを記憶するステップと、1次データレコードの更新を遠隔データプロセッサへ伝播するステップ(570)と、遠隔データプロセッサの各々にそのデータプロセッサに記憶された各2次コピーと対応する現行の満了時間を記憶するステップと、データプロセッサのデータアクセスの要求に応じて、現行の満了時間が過ぎたかどうかを検査するステップ(510)と、から成る。

1次コピーホルダ



【特許請求の範囲】

【請求項1】 分散データを処理する方法であって、少なくとも1つの遠隔データプロセッサ（200、250）において、第1データプロセッサ（100）で記憶される1次データレコードの複写された2次コピー（310）を記憶するステップと、前記1次データレコードの更新を前記遠隔データプロセッサ（200、250）へ伝播するステップ（420、570）と、前記遠隔データプロセッサ（200、250）の各々に前記遠隔データプロセッサに記憶される各前記2次コピーと対応する現行の満了時間（320）を記憶するステップと、前記遠隔データプロセッサのデータアクセスの要求に応じて前記現行の満了時間（320）が経過したかどうかを検査するステップ（510）と、から成る分散データ処理方法。

【請求項2】 前記現行の満了時間内に更新が前記1次コピーに行われるときに有効性インジケータ（330）を設定するステップ（460、590）と、前記遠隔データプロセッサにおけるデータアクセスの要求に応じて、前記遠隔データプロセッサによって保持される前記2次コピーが有効かどうかを決定するために前記有効性インジケータ（330）を検査するステップ（500）と、を含む請求項1に記載の分散データ処理方法。

【請求項3】 非有効性及び現行の満了時間の経過の内の一方を決定するときに2次コピーの更新を要求するステップ（530）を含む、請求項1又は請求項2に記載の分散データ処理方法。

【請求項4】 分散データを処理する装置であって、データレコード（310）の1次コピーを記憶するための手段を有する第1データプロセッサ（100）と、前記1次コピーの全て又は一部の2次コピーを記憶するための手段と、各前記2次コピーと対応する現行の満了時間を記憶するための手段と、前記現行の満了時間が経過したかどうかを検査するための手段と、を有する少なくとも1つの遠隔データプロセッサ（200）と、前記1次コピーへ行われる更新を前記2次コピーへ伝播する（420、570）ための手段と、から成る分散データ処理装置。

【請求項5】 更新が前記1次コピーに行われるときに有効性インジケータ（330）を設定するための手段と、前記遠隔データプロセッサ（200）におけるデータアクセスの要求に応じて前記有効性インジケータ（330）を検査する（500）ことによって、前記コピーが有効かどうかを決定するための手段と、を含む請求項4に記載の分散データ処理装置。

【請求項6】 非有効性及び現行の満了時間の経過の内の一方を決定するときに2次コピーの更新を要求する（530）ための手段を含む、請求項4又は請求項5に記載の分散データ処理装置。

【請求項7】 最後の更新が2次コピーを保持する前記

遠隔データプロセッサに伝播されて以来、1次コピーのどのデータレコードが更新されたかを決定する（400）ための手段と、前記遠隔データプロセッサへ前記更新されたデータレコードのコピーを伝播するための手段と、を有する請求項4乃至請求項6のいずれか1つに記載の分散データ処理装置。

【請求項8】 データレコードの更新されたコピーが前記1次コピーへアクセスするために前記遠隔データプロセッサ（200）からの要求のみで前記遠隔データプロセッサ（200）へ伝播され（570）、前記遠隔データプロセッサ（200）が前記更新されたコピーを記憶するための手段を有する、請求項4乃至請求項7のいずれか1つに記載の分散データ処理装置。

【請求項9】 1次データレコードが前記1次データレコードの異なる部分の前記1次コピーを保持するために異なるプロセッサに分散されるよう論理的に区分される、請求項4乃至請求項8のいずれか1つに記載の分散データ処理装置。

【請求項10】 時間可変の現行の期間を設定するための手段を含む、請求項4乃至請求項9のいずれか1つに記載の分散データ処理装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、分散データ処理システムに関する。

【0002】

【従来の技術】分散システムは、データ通信ネットワークによって相互接続される多くのデータ処理機械から成る。分散システムの重要なクラスは、1つのサイトにあるデータがもう1つのサイトで実行するデータ処理プログラムによって明らかにアクセス可能なことである。当該分散データベースシステムの一般的な説明は、デート著(C J Date)「分散データベースシステムとは何か(What is a Distributed Database System)」第1部、第2部(The Relational Journal, Nos 1 and 2, 1987年)に記載されている。

【0003】分散データベースシステムにおいて、通信ネットワークのデータ通信量を減少するため、データへアクセスするプロセスに近接してデータを配置する目的によって、データは分割されて幾つかのサイトに記憶されてもよい。しかしながら、通常、幾つかのサイトはもう1つのサイトに配置されたデータへアクセスしなければならない。この遠隔アクセスによってデータ処理操作に含まれるコスト及び遅延が増加し、これらのサイトの処理パフォーマンスはそれ自体のデータを持つ同等の独立型システムの処理パフォーマンスよりも極めて悪い。

【0004】他のネットワークのサイトにおいて通信リンク又はデータ処理機械が故障すると、遠隔データは一定の時間アクセスを妨げられるという追加の問題がある。従って、データの使用可能性は各サイトが独立型シ

システムであったならばより一層悪くなる。分散システムの目的はユーザがデータ資源を共用することだが、これらのマイナス効果はユーザが遠隔データアクセスに依存するのを妨げようとする。これはまた、単純な集中システムと比較すると分散システムの価値を下げている。

【0005】

【発明が解決しようとする課題】従って、分散システムのフィールドにおける不変の目的は、局所的な（ローカル）データに対して入手可能なパフォーマンス及び使用可能性と可能な限り一致するパフォーマンス及び使用可能性を備えた遠隔データへのアクセスを提供することである。この目的を達成するための1つの方法はネットワークを介してデータを複製することであり、殆どのデータアクセスが必要なデータの局所的なコピー又は近くのコピーによって満足可能である。この方法はサン(Sang Hyuk Son)著(SIGMOD Record, Vol 17 No 4 (1988))の文献に記載されている。この技術では、データアクセスのためのネットワーク通信量の減少と、データの複製のコピーへ更新を伝播するのに必要な追加のネットワーク通信量との間に均衡が見出されなければならない。

【0006】データ複製は、キャッシングを用いるローカルエリアネットワークフィアルサーバから、複製されたデータを用いる分散データベース管理システムに及ぶ分散システムの幾つかのタイプに使用されている。複製されたデータシステムの重要なクラスは、データオブジェクトの1次コピーが単一のデータプロセッサに保持され、データオブジェクトの他の全てのコピーが2次コピーとして指定される。更新のタイムシーケンスが正確なことを保証するために、先ず1次コピーが更新される。次に、改訂された1次コピーに基づいて2次コピーが改訂される。

【0007】データオブジェクトが高い割合の読み取りアクセスと、低い割合の書き込みアクセスとを有するとき、データオブジェクトの複製は最も有用である。これは、読み取りアクセスが単一の2次コピーで実行される一方、書き込みアクセスが1次コピーで実行されて、全ての2次コピーへ伝播されなければならないためである。従って、書き込みアクセスのコストは読み取りアクセスのコストより高い。分散システムでは、データオブジェクトの更新によって、データオブジェクトの遠隔2次コピーが無効にされ、ネットワークを介して伝送される新たなコピーによって置換されて、ネットワークのコストは他の更新のコストに追加されなければならない。

【0008】この方法の極端な例として、意思決定支援アプリケーションのために、読み取り専用の複製とされる「スナップショット(snapshots)」を使用する。リンゼイ(B. Lindsay)他著、「スナップショット差動リフレッシュアルゴリズム(Snapshot Differential Refresh Algorithm)」(IBM Research Report RJ4992, 1986年)では、スナップショットを1次データの現行の状態によ

り近く保つためにスナップショットがどのように定期的にリフレッシュ(再生)されるかが記載されている。しかしながら、スナップショットは保全性が保証されておらず、トランザクション中のデータを更新するために使用することができない。

【0009】多数のユーザが共用されたファイル又はデータベースを更新する場合、2次コピーがすばやく無効にされ、多くのネットワーク通信量が生成されてもよい。この追加の通信量は、複製がもたらす予定のネットワーク通信量の減少を越えることすらある。実際の結果として、シェール(A L Sherr)著、「ネットワークのシステムのための構造(Structures for Systems of Networks)」(IBM Systems Journal Vol 25, No 1, 1987年)に記載されるように、複製方法は、殆ど常に集中している巨大な共用されたファイル及びデータベースに有用ではない。

【0010】従って、従来技術における重要な問題は、データ複製が望ましくとも、データがユーザによって複製のサイトで更新可能であることが重要な場合にはデータ複製を達成するのが極めて困難なことである。しかしながら、多くの実際の場合では、分散されたサイトにおけるアプリケーションは最新のデータへのアクセスを必要とせず、公知の制御された量によって使い古されているデータで十分に機能することができる。この例として、定期的に更新される比率テーブルを使用するアプリケーション、及び現金の引き出しを許可するときに使い古された口座残高を用いる自動窓口システムがある。

【0011】

【発明が解決しようとする課題】従って、分散システムのフィールドにおける不変の目的は、局所的な（ローカル）データに対して入手可能なパフォーマンス及び使用可能性と可能な限り一致するパフォーマンス及び使用可能性を備えた遠隔データへのアクセスを提供することである。

【0012】

【課題を解決するための手段と作用】本発明に従って、分散データ処理装置及び分散データ処理方法が提供される。本方法は、少なくとも1つの遠隔データプロセッサに第1データプロセッサで記憶される1次データレコードの複製された2次コピーを記憶するステップと、前記1次コピーの更新を遠隔データプロセッサへ伝播するステップと、遠隔データプロセッサの各々にそのデータプロセッサに記憶された各2次コピーと対応する現行の満了時間を記憶するステップと、そのデータプロセッサにおけるデータアクセスの要求に応じて、現行の満了時間が過ぎたかどうかを検査するステップと、から成る。

【0013】好ましくは、肯定的又は否定的な有効性メッセージが遠隔プロセッサに記憶され、肯定的なメッセージは対応するコピーが有効でアクセスされてもよいことを示す一方、否定的なメッセージはコピーがアクセス

されないことを示す。後者の場合、データレコードの無効なコピーへアクセスを試みるアプリケーションプログラムは、1次データレコードを保持するプロセッサへ向けられる。

【0014】無効性が決定されるときにコピー全体がリフレッシュされる必要がなく、1次データレコードのどの部分が現行の期間内に更新されるかについてのレコードが保持され、他のプロセッサへ更新のみが伝播されるのが好ましい。

【0015】更に、全てのプロセッサへ全ての更新を伝送するのが不要になる。即ち、現行の期間内に遠隔プロセッサが1次レコード又は更新された部分さえにもアクセスを要求するならば、本発明の1態様に従って、1次データレコードに対する更新が遠隔プロセッサのみに伝播するシステムを提供するのが好ましい。

【0016】分散データを処理する装置は、データレコードの1次コピーを記憶するための手段を有する第1データプロセッサと、前記1次コピーの全て又は一部の2次コピーを記憶するための手段と、各前記2次コピーと対応する現行の満了時間を記憶するための手段と、前記現行の満了時間が経過したかどうかを検査するための手段と、を有する少なくとも1つの遠隔データプロセッサと、前記1次コピーへ行われる更新を前記2次コピーへ伝播するための手段と、から成る。

【0017】

【実施例】図1を参照すると、データ処理機械のネットワークの部分が示されて、データプロセッサ100、200及び250がデータ通信ネットワーク260によってリンクされている。データプロセッサ100、200は、中央処理装置(CPU)110、210から成り、対応するユーザ端末120、220及びデータ記憶装置130、230を備えている。データプロセッサはまた自動窓口機(ATM)又はプリンター等の追加の周辺装置を有してもよい。追加の装置は、概略的に140、240として示される。各プロセッサ100乃至250は、ライントランザクション処理アプリケーションを支援するために、CICS(顧客情報管理システム)として知られるデータベース/データ通信プログラムを実行する。CICSプログラムは、IBMの「CICS一般情報マニュアル(CICS General Information Manual)」(GC33-0155-3)に記載されている。

【0018】データネットワークによってデータプロセッサを互いにリンクすることは、IBMのマニュアル「CICS製品間通信(CICS Inter-Product Communication)」(SC33-0824-0)に記載されており、各データプロセッサはCICSシステム間通信機能を用いてもう1つのデータプロセッサに記憶されるデータへアクセスすることができる。例えば、CPU210によって必要とされる特定のデータレコードが実際、記憶装置130に記憶されるならば、CICSプログラムはネットワーク2

60を介してデータプロセッサ200からデータプロセッサ100へアクセス要求を送信して、要求されたデータレコードで応答する。この遠隔ファイルへアクセスする方法は、機能伝送として知られている。

【0019】概してネットワーク260等のデータネットワークは導入及び使用するのが高価なために、しばしばネットワーク通信量を可能な限り減らす努力がなされている。通信量の減少を達成するための1つの方法として、データファイルをネットワークの異なる機械の間で分割する方法があり、そのデータを最も頻繁に使用するデータプロセッサにて、或いは、そのデータプロセッサに物理的に近いところにデータが記憶される。しかしながら、同一のデータファイルが多数のデータプロセッサによって使用されるならば、ネットワーク通信量を減らすようにはならない。なぜならば、これら機械の内の1つを除く全てが、そのファイルがどこに保持されていようと、ネットワークを用いて単一のファイルへアクセスしなければならないからである。

【0020】従って、第2の方法はデータファイルの複数のコピーを作成して、ネットワークのさまざまなデータプロセッサにこれらのコピーを記憶することである。このようにして、特定のファイルのコピーが各データプロセッサで記憶されるか、或いは、少なくともこれらのファイルを通常用いるデータプロセッサで記憶される。各データプロセッサでファイルを複写することによって、データへのアクセスは局所的であり、ネットワークを介して伝送される必要のないことが保証される。

【0021】複写されるコピーの内の1つは、「マスター(master)」コピー又は「1次(primary)」コピーと称される。ファイルが更新されるときはいつも最初にマスターコピーが更新され、他の2次コピーに対する更新はマスターから取り出されて2次コピーへ伝播される。単一のマスターコピーを最初に更新することによって、正確なタイムシーケンスで更新が行われていることが保証される。全てのファイルのマスターコピーが単一のデータプロセッサによって保持される必要はない。特定のファイルのマスターコピーの場所は、そのファイルへ最もアクセスを必要とするデータプロセッサに、或いは、データプロセッサの近くに選択される。

【0022】マスターコピーが更新されるといつも、その更新されたレコード、或いは、更新されたファイルの完全なコピーさえも2次コピーの各々に伝播される。しかしながら、これによって、複写によってもたらされるいかなる利点も十分に否定することが可能な大量のネットワーク通信量が生じる可能性がある。本実施例において、この問題は2次コピーの各々に所定の不確実度を許容することによって処理される。これを達成するため、2次コピーの各々は対応する現行の期間を有し、現行の期間の満了時間がそのファイルに対応するCICS制御ファイルに記憶される。マスターコピーが平均時間で更

新されたかどうかにかかわらず、現行の満了時間が過ぎるまで、2次コピーは有効であると仮定される。2次コピーが1次コピーの全てのレコードを含む必要はなく、2次サイトで頻繁に必要とされるレコードだけを含んでもよい。従って、異なる2次コピーは異なるレコードを含んでもよい。

【0023】図2は、2次コピーを構成する記憶されたデータレコードとの現行の満了時間の対応を示す概略図である。図2において、データレコードはCICSデータファイル300に配置されるように示されている。ファイル300の各レコードに対応して、現行の満了時間(T)320及び有効性フラグ(V)330があり、ファイル制御テーブル340に記憶される。現行の満了時間の起点は以下に記載される。データレコード310を読み取ろうとすると、データレコード310に対応する有効性フラグ330が最初に検査される。有効性フラグ330が「無効(invalid)」と設定されなければ、現行の満了時間320が経過したかどうかを決定するために現行の時間が検査される。現行の満了時間320が経過していなければ、データレコード310が読み取られ、動作は通常どおり続く。現行の満了時間320が経過したならば、レコードは読み取られず、そのレコードに対応する有効性フラグ330が「無効(invalid)」と設定され、2次コピーの更新が要求される。追加の動作は図4にも示される。

【0024】現行の満了時間Tは、個々のデータレコード又は全体のファイルに対応する。データレコードへの更新が伝播されるときはいつもマスターコピーのホルダによって伝播される。また、マスターから更新が受信されるときはいつも、マスター更新が発生した時間に予め定められた現行の期間を追加することで、現行の満了時間は2次コピーのホルダによって再設定される。

【0025】データレコードの実際の現行の期間（現行の満了時間Tとマスターレコードが更新される時間との差）は、ユーザによって、データレコードの性質、及び任意ではあるが、現行の時間又は現行の日付に関連して決定される。しかしながら、主にデータの性質に依存している。一例として、分散システムに記憶されるデータが銀行の口座残高に関連する場合がある。（例えば、自動窓口機(ATM)を用いて）少額の現金の引出しを許可する目的のために、ユーザとしての銀行は、局所的な2次コピーで保持される残高を、例えば、24時間毎より頻繁に更新するのに、余分なネットワークコスト及び処理コストを払う価値はないと考えるかもしれない。

【0026】上記のように、現行の期間は現行の時間又は日付によって変更可能である。例えば、更新がその時間に特に頻繁に行われるならば、現行の期間は月曜日の朝（営業週間の始まり）により一層短くする必要があるかもしれない。同様に、現行の期間は殆どデータ更新動

作のない週末に延長することも可能である。

【0027】図3は、少なくとも1つのデータレコードの1次コピーから2次コピーへの更新を伝播するための方法を示すフローチャートである。「ニュース速報(news bulletin)」方法又は「プッシュダウン(pushdown)」方法と称されるこの方法は、2次コピーの現行の満了時間を知っているマスターコピーのホルダに依存するが、いかなる読み取り要求が行われるかどうかには依存する必要はない。

10 【0028】フローチャートに示される幾つかのステップは1次コピーのホルダによって実行され、残りは2次コピーのホルダの各々によって実行される。1次コピーホルダと2次コピーホルダの間の通信は、CICSシステム間通信機能、或いは、信頼できるメッセージ機能等の任意の同等の機能を介して達成される。

20 【0029】プロセスはステップ400で開始し、最後の更新が2次コピーへ伝播されて以来、1次コピーホルダは1次コピーのどのデータレコードが更新されたかを決定する。本実施例において、ステップ400は2次コピーで保持される現行の満了時間に達する前に実行される。ステップ410において、新たな現行の満了時間Tは、現行の時間及びデータレコードの性質に関連する所定の情報に基づいてデータレコードのセットに対して計算される。この現行の満了時間のレコードは、1次コピーホルダによって保持される。

30 【0030】ステップ420において、全体のファイルと対応する現行の満了時間と共に、更新されたデータレコードは2次コピーホルダへ伝播される。2次コピーホルダによって実行される第1ステップは、ステップ430において更新を受信し、ファイルの2次コピーの更新を行うことである。ステップ450において、データレコードに対応する新たな現行の満了時間が記憶される。最後に、これらレコードと対応するあらゆる有効性フラグが、データレコードの有効性期間が満了したと示す「無効(invalid)」と設定されるならば、有効性フラグはステップ460において再設定される。この方法は、現行の期間の満了の前に更新が2次コピーへ伝播されるならば、各2次コピーは常に使用できることを保証する。

40 【0031】図4は、ファイルの1次コピーから2次コピーへ更新を伝播するための第2の方法を示すフローチャートである。この方法は、「即時回答(on demand)」方法又は「プルダウン(pulldown)」方法と称される。また、フローチャートで詳細に示される幾つかのステップは1次コピーホルダによって実行され、残りは2次コピーホルダの各々によって実行される。

50 【0032】図4に示されるプロセスは、2次コピーホルダが2次コピーに保持されるデータレコードの読み取り動作を開始するときを開始する。データレコードに対応する有効性フラグは、ステップ500で検査される。

フラグが「無効(invalid)」と設定されるならば、制御は直接ステップ530へ渡されて、そのデータレコードに対する更新が要求される。これらの場合では、有効性フラグを使用することによってデータレコードの有効性を更に検査する必要性が取り除かれる。しかしながら、有効性フラグが「有効(valid)」と設定されるならば、ステップ510において、現行の時間がデータレコードと対応する現行の満了時間よりも遅いかどうかを決定するための追加の検査が行われる。現行の時間がTよりも早いならば、データレコードはステップ540において局所的な2次コピーから読み取られて、プロセスは無事に終了する。

【0033】ステップ510において、現行の満了時間が経過したならば、制御はステップ520に渡されて有効性フラグが「無効(invalid)」と設定され、ステップ530に渡されてそのデータレコードに対する更新が1次コピーホルダから要求される。しかしながら、この要求が合わない場合がある。例えば、1次コピーホルダが故障した、或いは、通信ネットワークが利用不可能となった場合である。ステップ550において更新要求が与えられると、制御は1次コピーホルダへ渡され、ステップ560において要求されたレコードを検索しようと試みる。レコードが発見されると、1次コピーホルダは新たな現行の満了時間を計算し、ステップ570において更新されたレコード及び新たな時間の双方を2次コピーホルダへと伝送する。1次データレコードが最後に2次コピーホルダへ送信されて以来、1次データレコードが変更されていなかったならば、肯定信号及び新たな現行の完了時間が代わりに送信されてもよい。

【0034】ステップ550において、更新要求が与えられなければ、プロセスは不成功に終わる。この場合(データ利用不可)は、データレコードを読み取るための元の要求を作成したプログラムによって従来の方法で処理される。

【0035】ステップ580において、更新は2次コピーホルダによって受信されて2次コピーへ適用される。ステップ590において新たな現行の満了時間Tが保管(セーブ)され、データレコードと対応する有効性フラグVは「有効(valid)」と再設定される。次にデータレコードの局所的なコピーがステップ540で読み取られ、プロセスは無事に終了する。この方法はアプリケーションによって特別に要求されたときのみデータレコードを伝送することでネットワーク通信量を最小限にし、ときどき使用不可能なこれらのレコードを犠牲にして、有効なコピーは局所的には存在しない。

【0036】「即時回答(on demand)」方法及び「ニュース速報(news bulletin)」方法の特徴を組み合わせる、組合せ更新方法を用いることは可能である。この場合、遠隔サイトで保持される2次コピーは、いかなるデータも保持しない空のファイルとして開始する。次に、

これは「即時回答(on demand)」方法を用いる遠隔サイトの必要性に適したデータレコードのセットへと組み込まれる。いったんデータレコードの作業セットが設定されると(「即時回答(on demand)」方法が適切な時間で作動すると)、動作は「ニュース速報(news bulletin)」方法へ切り替えることが可能である。また双方の方法は同時に作動可能であり、その場合には遠隔サイトで保持される2次コピーは、遠隔サイトによって必要とされたデータレコード及び1次コピーで最近更新されたデータレコードから成る。組合せの方法は2次コピーが作成されるにつれて追加の処理を要するが、その後は「ニュース速報(news bulletin)」方法と同様に実行する。組合せの方法は、2次コピーがいかなる理由で失われても自動的に回復されるという利点がある。

【0037】分散複写システムのデータレコードの更新が行われると、データレコードのあらゆる他のコピーが更新される前に1次コピーが更新される。これは、更新が正確なタイムシーケンスで行われることを保証するためである。しかしながら、多くの分散データ処理システム(トランザクション処理システム等)において、データは遠隔サイトで捕獲されて、1次コピーへ転送される。単一のサイトが全てのデータの1次コピーホルダならば、これらのデータ捕獲動作で1次コピーを更新しなくてはならないために、それらのパフォーマンスは非複写システムの場合と同じである。更新フローは図5に示され、ここで遠隔データプロセッサは、現行のCICSシステム間通信機能を用いて1次コピーファイルを更新する。ステップ600において、遠隔データプロセッサは更新のためのレコードを用意し、ステップ610において、中央データプロセッサへそのレコードを伝送する。ステップ620においてこのレコードが受信され、ファイルの1次コピーが更新される。ステップ630において更新の確認が生成されて遠隔サイトへ伝送され、ステップ640で受信される。なお、ネットワーク故障が発生する場合には、ステップ610において更新は完全に失敗する。

【0038】これらパフォーマンス問題及び利用可能性問題に対する1つの解決法が図6に示され、ここで1次コピー自体がネットワークを介して分散される。換言すれば、各サイトは、可能な限り更新を一般的に必要とするそれらデータレコードの1次コピーホルダである。中央サイトがそれらデータレコードの2次コピーを保持することで、完全なファイルは他のサイトへの複写又は中央処理のために利用することができる。

【0039】図6は、図1に示される分散データ処理システムと同様の分散データ処理システムを示している。データ記憶装置130、230に記憶されるデータは、4つの区画A、B、C及びDから成るとして概略的に示されている。中央側にあるデータプロセッサ100は区画Cの1次コピー670を保持し、データプロセッサ2

00は区画Bの1次コピー680を保持する。各データプロセッサは、そのデータプロセッサのデータアクセスの必要性に適するならば他の区画の2次コピーを保持してもよい。区画は重なり合わないよう画定されて、いかなる所定のレコードにも1つの1次コピーしかないことが重要である。これは、レコードをキー範囲又はあらゆる同様の方法で区分することによって達成可能である。概して、各区画は対応する記述子650、660を有して、データプロセッサに特定のレコードがこの区画にあるかどうかを決定させる。

【0040】改訂された更新フローは図7に示され、ここで遠隔データプロセッサがレコードを更新する。ステップ700において、遠隔データプロセッサは更新するためのレコードを用意し、ステップ710において、局所的な区画記述子をテストして、レコードがそのサイトで保持される1次コピー区画内に含まれるかどうかを決定する。含まれるならば、ステップ720においてこの区画に更新が行われ、ステップ730において更新の確認が生成される。更新されたレコードが局所的な区画内に含まれなければ、ステップ740において中央サイトへ伝送され、ステップ750において受信され、中央ファイル（及び異なるならば1次コピー）へ適用される。ステップ760において、確認が生成され、遠隔サイトへ伝送される。局所的な区画がそのサイトで更新されるレコードの大部分を含むように選ばれるならば、この方法はネットワーク通信量をかなり減少することができる。この方法はまた、中央サイトへ伝送されなければならない更新だけがネットワーク故障のために失敗するため、利用可能性も増やす。

【0041】ある場合において、特定の遠隔サイトはデータレコードのセットのために1次コピーホルダとして開始されてもよいが、異なる遠隔サイトは、その後多数の更新をそれらデータレコードの1次コピーへ行う必要がある。データ処理サイトの更新要求において、この変化によって生成されるネットワーク故障及び追加のネットワーク通信量による分裂の可能性は、分散1次コピーを用いることによって達成される利点の幾つかを否定するかもしれない。この問題に対する解決法は、1次コピーを保持する責任をサイト間で動的に転送するのを可能にすることである。このような構成は「チェックアウト／チェックイン(checkout/checkin)」方法と称され、図8に関連して説明される。

【0042】ステップ800において、遠隔サイトにおけるデータプロセッサは、中央サイトに区画記述子を伝送することによって、ファイルの特定の区画に責任を要求する。ステップ810でこの区画が検索され、ステップ820において、制御ファイルに保持された既にチェックアウトされた区画の記述子と比較される。要求された区画がもう1つのサイトへとチェックアウトされたか、或いは、既にチェックアウトされた区画と重なり合

うならば、要求は失敗する。ステップ830において、エラーメッセージが生成され、要求サイトへ伝送される。ステップ840において、このメッセージが受信されて、プロセスが不成功に終わる。しかしながら、要求された区画がチェックアウトされていなかったならば、その記述子はステップ850において制御ファイルに保管される。確認メッセージがステップ860において作成され、1次コピーホルダとなる遠隔サイトへ伝送される。このメッセージはステップ870で受信され、プロセスは無事に終了する。その後、遠隔データプロセッサはこの区画の制御を解放するために同様のプロセスを経て、中央サイトへ「それをチェックインする(checking it in)」。

【0043】1次コピーの責任の動的な転送は、上記の組合せ更新方法へと組み込まれる。データプロセッサのネットワークの動作の開始に際して、全てのデータの1次コピーは中央サイトと称される単一のサイトで保持される。処理が行われるにつれて、データレコードのコピーは「ニュース速報(news bulletin)」処理又は「即時回答(on demand)」処理によってネットワークの回りに分散される。しばらく後、分散サイトは特定の区画の所有権を要求し、この区画の1次コピーの責任はその該当するサイトへ動的に転送されてもよい。その後、ネットワークの負担が時間と共に変化するにつれて、責任は中央サイトへ戻されるか、或いは、もう1つのサイトへ転送されてもよい。

【0044】

【発明の効果】本発明は上記より構成され、局所的なデータに対して入手可能なパフォーマンス及び使用可能性と可能な限り一致するパフォーマンス及び使用可能性を備えた、遠隔データへのアクセスを提供する。

【図面の簡単な説明】

【図1】本発明が実施される分散データ処理システムの部分を示す。

【図2】本発明に従って、図1のシステムで使用されるデータ及び関連する制御テーブルを概略的に示す。

【図3】本発明に従って、分散データ処理システムで更新を伝播するための「ニュース速報(news bulletin)」又は「プッシュダウン(pushdown)」方法を示すフローチャートである。

【図4】本発明に従って、分散データ処理システムで更新を伝播するための「即時回答(on demand)」又は「プルダウン(pulldown)」方法を示すフローチャートである。

【図5】中央1次コピーへの更新を行うための従来の方法を示す。

【図6】ファイルが分散1次コピーを有する分散データ処理システムの部分を示す。

【図7】遠隔プロセッサが更新を行う図6のシステムで使用される更新方法を示すフローチャートである。

13

【図8】図6のシステムにおいて中央サイトから遠隔サイトへのデータレコードの1次コピーの責任の転送を示すフローチャートである。

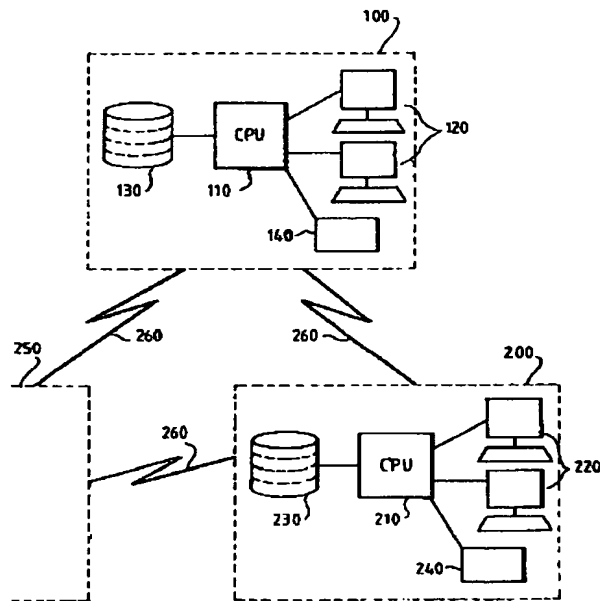
【符号の説明】

100、200、250 データプロセッサ

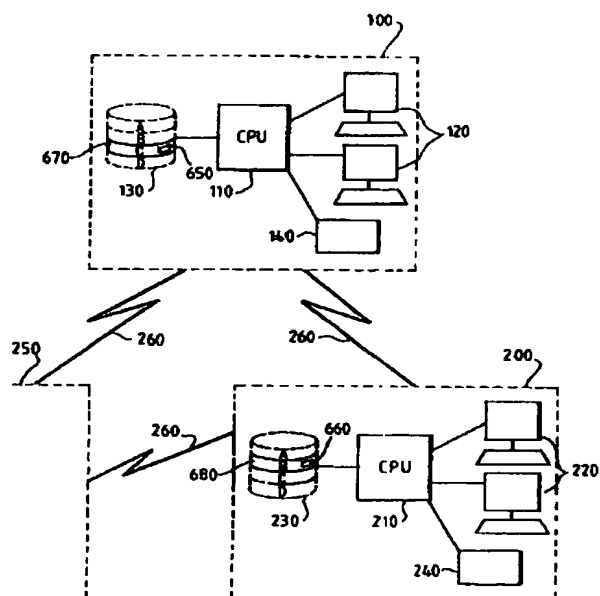
110、210 CPU

120、220 ユーザ端末

【図1】



【図6】



14

* 130、230 データ記憶装置

260 データ通信ネットワーク

300 データファイル

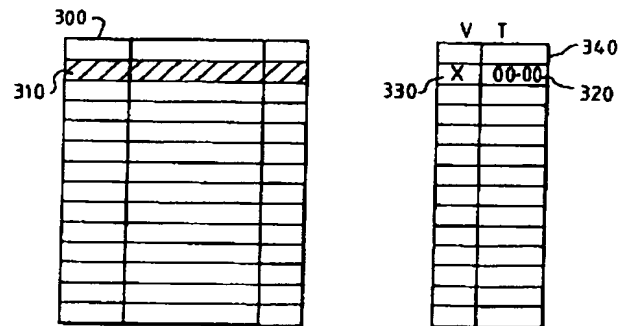
310 データレコード

320 現行の満了時間

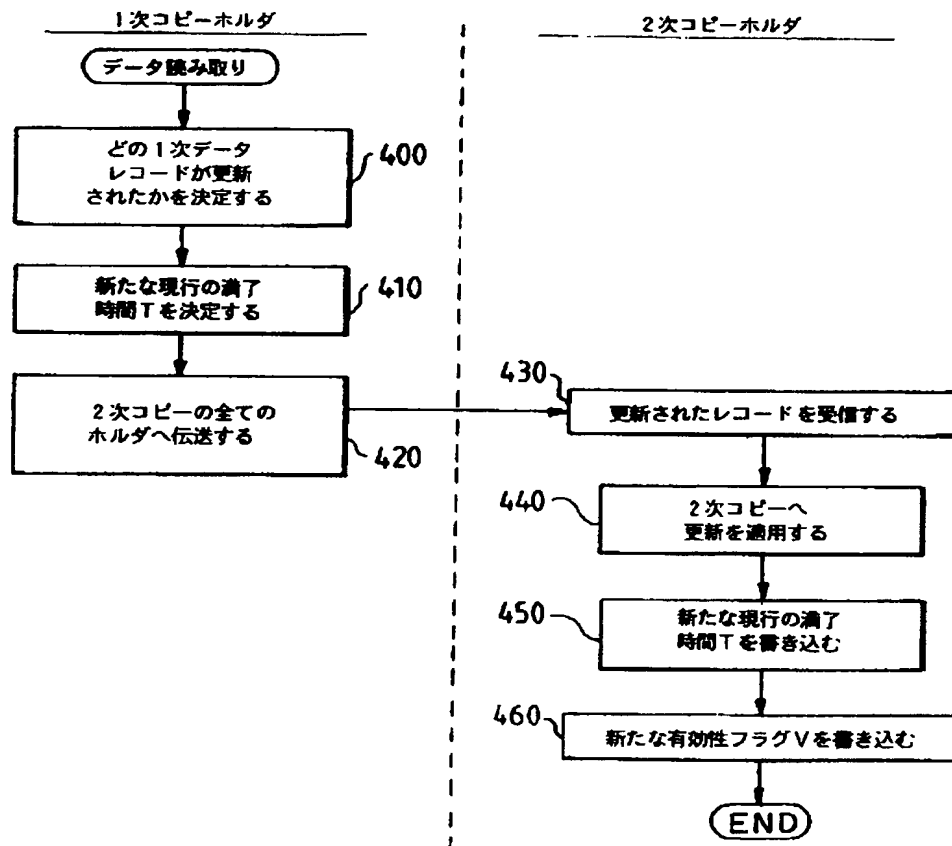
330 有効性フラグ

* 340 ファイル制御テーブル

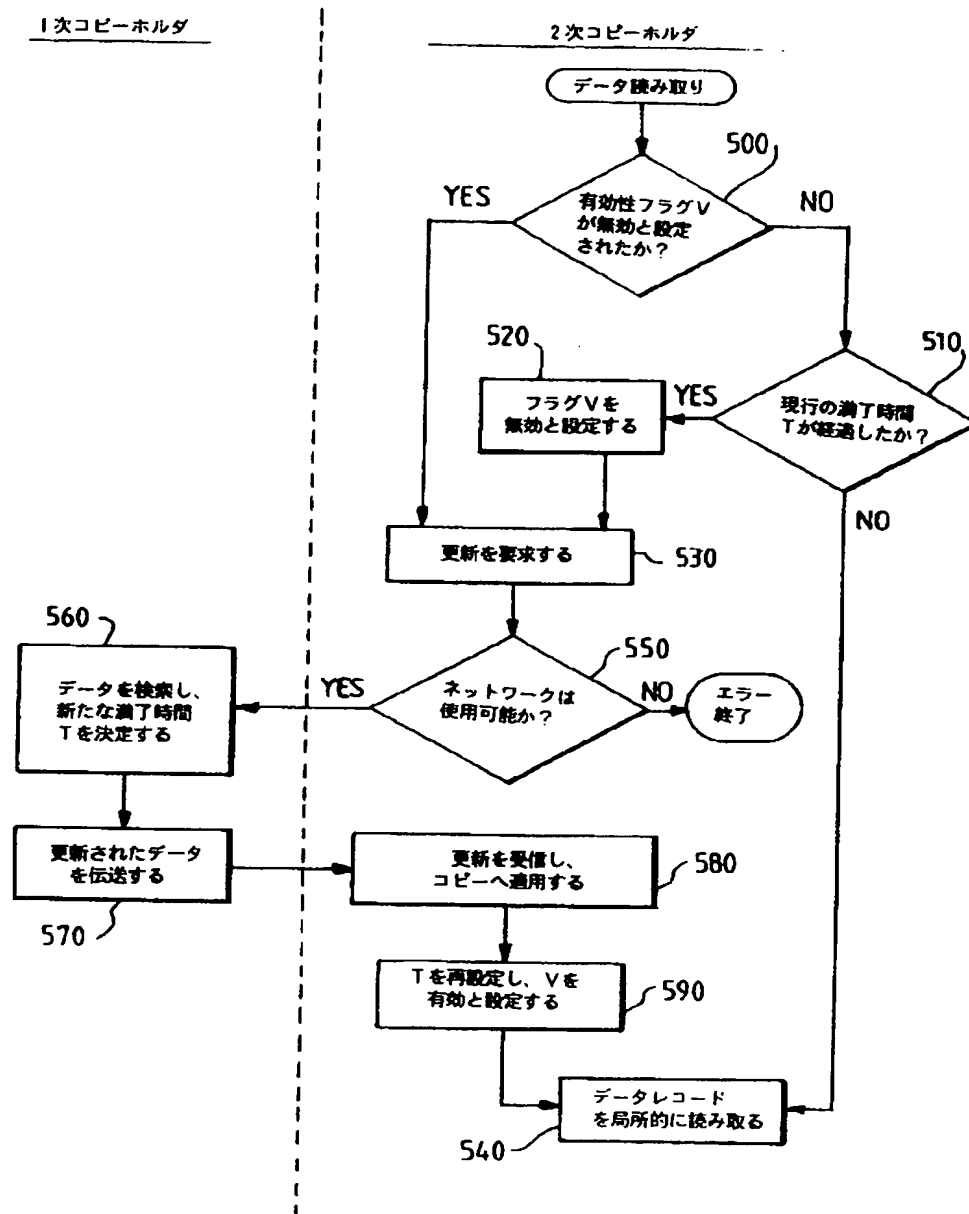
【図2】



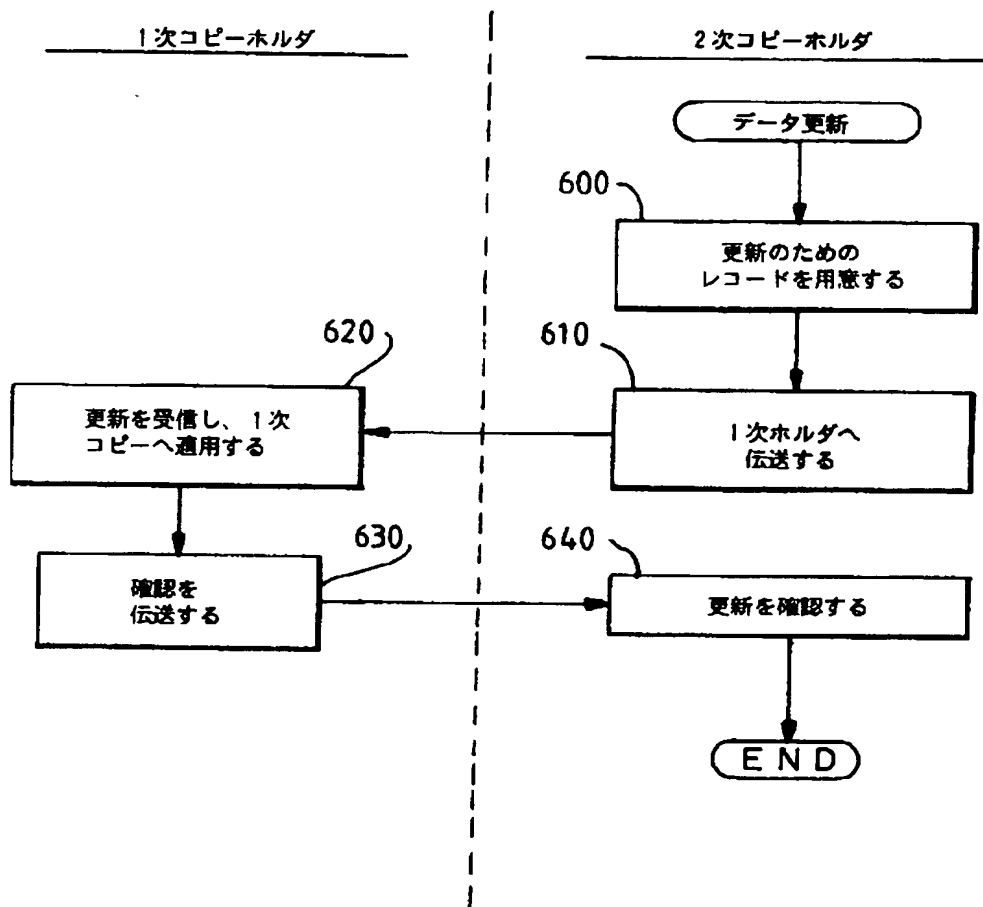
【図3】



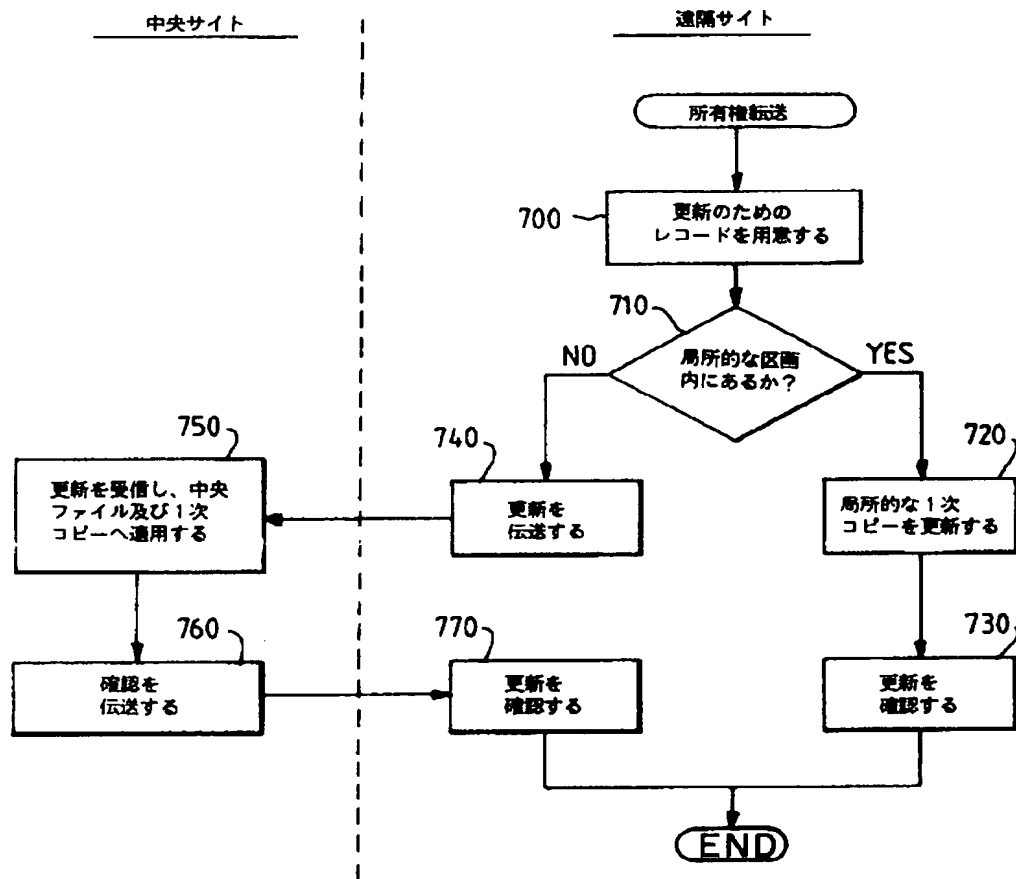
【図4】



【図5】



【図7】



【図8】

